

# Connecting Galaxies with Halos Across Cosmic Time: Stellar mass assembly distribution modeling of galaxy statistics

M. R. Becker<sup>1,2</sup>

<sup>1</sup>KIPAC, Physics Department, Stanford University, Stanford, CA 94305.

<sup>2</sup>KIPAC, SLAC National Accelerator Laboratory, Menlo Park, CA 94025.

15 July 2015

## ABSTRACT

In this work, I explore an empirically motivated model for investigating the relationship between galaxy stellar masses, star formation rates and their halo masses and mass accretion histories. The core statistical quantity in this model is the stellar mass assembly distribution,  $P(dM_*/dt|\mathbf{X}, a)$ , which specifies the probability density distribution of stellar mass assembly rates given a set of halo properties  $\mathbf{X}$  and epoch  $a$ . Predictions from this model are obtained by integrating the stellar mass assembly distribution (SMAD) over halo merger trees, easily obtained from modern, high-resolution  $N$ -body simulations. Further properties of the galaxies hosted by the halos can be obtained by post-processing the stellar mass assembly histories with stellar population synthesis models. In my particular example implementation of this model, I use the Behroozi et al. (2013a) constraint on the median stellar mass assembly rates of halos as a function of their mass and redshift to construct an example parameterization of  $P(dM_*/dt|\mathbf{X}, a)$ . This SMAD is then integrated over individual halo mass accretion histories from  $N$ -body merger trees starting at  $z = 4$ , using simple rules to account for merging halos. I find that this a simple model can reproduce qualitatively the bimodal features of the low-redshift galaxy population, including the qualitative split in the two-point clustering as a function of specific star formation rate. These results indicate that models which directly couple halo and galaxy growth through simple efficiency functions can naturally predict the star formation rate bimodality in higher-order statistics of the galaxy field, such as its two-point correlations or galactic conformity signals.

**Key words:**

## 1 INTRODUCTION

The connection between the growth of galaxies and dark matter halos in the cosmological context of a  $\Lambda$ CDM universe is one of the primary concerns of galaxy formation theory. In the modern era, a variety of approaches have been taken to study this important problem. These range from *ab initio* computations of the formation of galaxies in large, supercomputer simulations (see, e.g., Somerville & Davé 2014 for a review) to fitting empirical models to observations of galaxies in order to infer the properties of their host dark matter halos (i.e., the halo occupation distribution, HOD, or conditional luminosity function, CLF, models; see, e.g., Berlind & Weinberg 2002; van den Bosch et al. 2007; Zehavi et al. 2011; Cacciato et al. 2013). Between these two extremes, other approaches combine directly high-resolution, dark-matter-only supercomputer simulations of

structure formation with simplified prescriptions for galaxy formation. These approaches include semi-analytic models (SAMs) where one attempts to parameterize directly the physics of galaxy formation (e.g., White & Frenk 1991; Cole et al. 1994; Somerville & Primack 1999, see Baugh 2006 or Somerville & Davé 2014 for a review), or various forms of abundance matching, like subhalo abundance matching (SHAM, e.g., Kravtsov et al. 2004; Conroy et al. 2006; Conroy & Wechsler 2009; Reddick et al. 2013) and conditional abundance matching (CAM, Hearin & Watson 2013; Hearin et al. 2014; Watson et al. 2015), that take the empirically *anzatz* that the most massive galaxies occupy the largest halos on average. Yet another approach along these lines combines observations at multiple epochs with flexible models of the stellar mass-to-halo mass relation in order to constrain the growth of stellar mass in halos across cosmic time (e.g.,

Conroy & Wechsler 2009; Moster et al. 2010; Behroozi et al. 2013a).

None of these approaches alone is currently sufficient to fully describe the wealth of observations of galaxies. A key advantage of models like HOD/CLFs, SHAM/CAMs or models similar to that in Behroozi et al. (2013a), is that they are formulated empirically, allowing for data to directly constrain how galaxies occupy halos, and when and where dark matter halos form galaxies. Alternatively, *ab initio* simulations of galaxy formation offer the opportunity to directly test models of the primary physical processes that govern galaxy formation. SAMs can offer these insights as well.

In this preliminary work, I explore a statistical approach to modeling the galaxy-halo connection across cosmic time through the stellar mass assembly distribution (SMAD)

$$P(dM_*/dt|\mathbf{X}, a). \quad (1)$$

This probability density function describes the distribution of stellar mass assembly rates given a set of halo properties  $\mathbf{X}$  and epoch  $a$ . It is important to note, as described further below, that many previous authors have built models very similar to that proposed in this work (e.g., Wang et al. 2007; Mutch et al. 2013; Lilly et al. 2013; Lu et al. 2014, 2015). In this work, I propose to assume a form for the SMAD and then directly integrate it over the ensemble of halo mass accretion histories from high-resolution  $N$ -body simulations.

For simplicity and practicality, I assume the Behroozi et al. (2013a) relationship between halo mass and stellar mass assembly rate without any modifications. This relationship is used to build an example form of the SMAD, which is then applied directly to individual halo mass accretion histories from a high-resolution  $N$ -body simulation. Note that the Behroozi et al. (2013a) relation was originally constrained for median median mass accretion histories as opposed to individual mass accretion histories. I then perform some preliminary comparisons to local data, mostly from the Sloan Digital Sky Survey (SDSS, York et al. 2000; Abazajian et al. 2009). I find that while my model does not perfectly reproduce statistics from local data, it captures many of the qualitative features of local data, including the specific star formation rate (sSFR) bimodality in the clustering of galaxies. This result, in agreement with the physical interpretation of galactic conformity through the effects of tidal forces on the mass accretion rate of dark matter halos (Hearin et al. 2015), indicates that modeling galaxy stellar mass assembly in a way which is directly coupled to halo mass assembly can naturally predict the bimodal higher-order statistics of the galaxy density field. Future work exploring general parameterizations of the SMAD, directly fit to observational data, will improve the results presented here, but is well beyond the scope of this work.

This work is organized as follows. In Section 2, I describe the simulation and observational data used in this work. In Section 3, I further discuss the general principle of SMAD modeling and describe my example model. Section 4 compares the results of my example SMAD to the data from the local Universe, mostly the SDSS. Finally, I conclude in Section 5.

## 2 SIMULATION AND SDSS DATA

In this work, I use a  $250 h^{-1}\text{Mpc}$  box with  $2048^3$  particles, **c250**. The **c250** box has a flat  $\Lambda\text{CDM}$  cosmology with  $\Omega_m, \Omega_b, \sigma_8, n_s, h = \{0.286, 0.047, 0.82, 0.96, 0.7\}$  with  $h \equiv H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . The box was run with **LGadget-2**, an  $N$ -body only version of the **Gadget-2** code Springel (2005). The initial conditions were generated with **CAMB** (Lewis & Bridle 2002) and second-order Lagrangian perturbation theory at  $z = 99$  with the **2LPTic** (Crocce et al. 2006). One hundred snapshots were saved between  $z \approx 12$  and  $z = 0$ , evenly spaced in the logarithm of the scale factor. Halos and merger trees were constructed with the **ROCKSTAR** (Behroozi et al. 2013b) and **CONSISTENT TREES** codes (Behroozi et al. 2013c). All halos with at least 20 particles are used in integrating the stellar mass assembly rates over the merger trees. However, when comparing to the low redshift SDSS data, only halos with maximum circular velocity greater than  $55 \text{ km s}^{-1}$  are used.

I use the measurements of the projected correlation function of SDSS galaxies from Watson et al. (2015). These measurements were performed on the SDSS DR7 catalog (York et al. 2000; Abazajian et al. 2009) using all galaxies with stellar masses satisfying  $\log_{10}(M_*/M_\odot) > 9.8$ . The galaxies were split according to their star formation rates, computed using a combination of emission line indicators (i.e.,  $\text{H}\alpha$  Brinchmann et al. 2004; Salim et al. 2007) and  $\text{Dn4000}$  (Kauffmann et al. 2003) in the cases of no emission lines or AGN contamination. The catalog of star formation rates is publicly available.<sup>1</sup> Finally, for some comparisons to data, I use directly the SDSS mock constructed by Watson et al. (2015). This mock catalog successfully predicts a large number of measurements from the SDSS, include the projected correlation function as a function of sSFR, galaxy-galaxy lensing as a function of sSFR, the colors and sSFR's of galaxies in galaxy groups, and the colors and sSFR's of central galaxies (Hearin & Watson 2013; Hearin et al. 2014; Watson et al. 2015).

## 3 CONNECTING GALAXIES WITH HALOS ACROSS COSMIC TIME

In this work, I explore how one can use draws of merger trees from  $N$ -body simulations to perform the integral of the SMAD over the population of halos in  $\Lambda\text{CDM}$ -like universes.<sup>2</sup> This approach has several important features.

(i) The SMAD is statistical in nature, aiming to predict (or constrain!) the ensemble of stellar mass assembly rates (and thus galaxy properties) from the ensemble of halo properties. Note that any purely deterministic model is the limit  $P(dM_*/dt|\mathbf{X}, a) = \delta(dM_*/dt - f(\mathbf{X}, a))$  where  $f(\mathbf{X}, a)$  is some function specifying the stellar mass assembly rate.

(ii) The statistical nature of the SMAD means that one does not have to specify all (or any, see the next feature) of the physical processes which control star formation for any individual halo (and information about which may not

<sup>1</sup> <http://www.mpa-garching.mpg.de/SDSS/DR7>.

<sup>2</sup> Analytical integrals of this PDF over stochastic processes which model the mass accretion histories of halos may be a useful avenue for future work, but are neglected in this work.

be present in a dark matter-only simulation). Instead, the SMAD posits that on average halos of a given set of properties form stars at a given rate and that the scatter about this average is a property of the ensemble of halos, characterizing something intrinsic to galaxy formation itself.

(iii) One can take a largely physics-agnostic approach by fitting  $P(dM_*/dt|\mathbf{X}, a)$  to observational data in the spirit of works which combine abundance matching or direct models of the stellar mass-to-halo mass relation across epochs (e.g., Conroy & Wechsler 2009; Moster et al. 2010; Behroozi et al. 2013a), SHAM/CAM models (e.g., Kravtsov et al. 2004; Conroy et al. 2006; Hearin & Watson 2013; Hearin et al. 2014; Watson et al. 2015) and HOD/CLF models (e.g., Berlind & Weinberg 2002; Zehavi et al. 2011; Cacciato et al. 2013).

(iv) Unlike other common statistical and empirical descriptions of the galaxy population, the SMAD explicitly connects galaxies across cosmic time for individual halos. In particular, integrating the SMAD over the merger history of dark matter halos assigns a star formation history to every halo in the simulation in an empirical manner. Thus it is possible to directly predict the colors/SEDs of galaxies, the rate of SNe as a function of epoch, the overall star formation rate density of the Universe as a function of epoch, etc. Thus SMAD modeling will be a key tool for connecting observations of galaxies across cosmic time.

Note that this approach encapsulates that taken by many other authors (Wang et al. 2007; Mutch et al. 2013; Lilly et al. 2013; Lu et al. 2014, 2015) where deterministic, empirical parameterizations of the growth of galaxies are used with merger trees to predict galaxy properties.

Furthermore, the SMAD model is similar to SAMs, except for its largely empirical nature. In fact, SMAD models are properly thought of as hybrids between HOD/CLF or other empirical models and SAMs, retaining the empirical and statistical nature of these models while explicitly connecting halos across cosmic time like a SAM. SMAD models also have important relations to SHAM/CAM models. In particular, below I assume that  $dM_*/dt \propto dM_{\text{vir},\text{cen}}/dt$ . Thus any assembly bias (e.g. Sheth & Tormen 2004; Gao et al. 2005; Wechsler et al. 2006; Harker et al. 2006; Bett et al. 2007) contained in mass accretion histories of dark matter halos will be captured in the star formation histories, star formation rates and stellar masses of halos. To the extent that SHAM/CAM models prefer proxies which themselves capture assembly bias, like  $v_{\text{max}}$  or  $v_{\text{peak}}$ , one expects qualitatively similar predictions from the SMAD at any epoch to these models. This expectation is borne out below. Thus SMAD models can naturally incorporate assembly bias and could even constrain the degree to which stochasticity in the process of galaxy formation could erase assembly bias signatures present in the halo population.

### 3.1 An Example SMAD Model

The specific SMAD model in this work combines the results of Behroozi et al. (2013a) with the mass assembly history of halos from high-resolution  $N$ -body simulations to predict  $dM_*(\mathbf{X}, a)/dt$  and the full stellar mass assembly histories of the halos. Schematically, at the initial epoch stellar mass is seeded into the halos using a fiducial stellar mass-to-halo

mass relation. Then the amount of stellar mass assembled by each halo is drawn from the SMAD given its properties  $\mathbf{X}$  and the epoch  $a$ . Next, a simple set of rules is used to compute the stellar mass of the descendent halos from the progenitors in the tree. Finally, this process is repeated until the final epoch is reached.

#### 3.1.1 Assembling Stellar Mass with the SMAD

The SMAD model I use in this work is given by drawing  $dM_*/dt$  from

$$dM_*/dt \sim \text{Logn}(10^{-12} \text{yr}^{-1} M_*, \epsilon) + dM_*/dt|_{\text{B13}} \quad (2)$$

where

$$dM_*/dt|_{\text{B13}} = dM_*(M_{\text{vir},\text{peak}}, a)/dM_{\text{vir}}|_{\text{B13}} \times \begin{cases} dM_{\text{vir},\text{cen}}/dt|_i & dM_{\text{vir},\text{cen}}/dt|_i > 0 \\ 0 & dM_{\text{vir},\text{cen}}/dt|_i \leq 0 \end{cases},$$

$\epsilon = 0.25$ , and  $\text{Logn}(\mu, \sigma)$  is a log-normal distribution of mean  $\log_{10} \mu$  and scatter  $\sigma$ . I use base-10 logarithms in this work.  $dM_*(M_{\text{vir},\text{peak}}, a)/dM_{\text{vir}}|_{\text{B13}}$  is the Behroozi et al. (2013a) result and  $dM_{\text{vir},\text{cen}}/dt|_i$  is the mass accretion history of halo  $i$  in the simulation, as computed below.  $M_{\text{vir},\text{peak}}(a)$  is the peak mass along the most massive progenitor branch of the merger tree of each halo and is the quantity used in Behroozi et al. (2013a) in order to define their parameterization of stellar mass assembly rates. The first term in this model is a rough, *ad hoc* way to account for residual star formation in quenched galaxies, information about which is not present in dark-matter-only  $N$ -body simulations. Note that because of this choice, the quenched peak in the distribution of sSFR's in the mock is placed at the correct location and given roughly the correct width “by hand.” Furthermore, this model assumes that no dark matter stripping ever corresponds to the stripping of stellar mass. Such an assumption may not be warranted, but I leave exploring this issue to future work. Finally, note that the pseudo-evolution of  $M_{\text{vir}}$  (i.e., changes in  $M_{\text{vir}}$  due to evolution in the over density used to define the halo as opposed to accretion of the mass onto the halo itself, see Diemer et al. 2013) cancels in the computation of the stellar mass assembly rates.

The mass accretion rate  $dM_{\text{vir},\text{cen}}/dt$  is needed in order to compute  $dM_*/dt$  above. For a halo that is not merging with another at a given epoch,

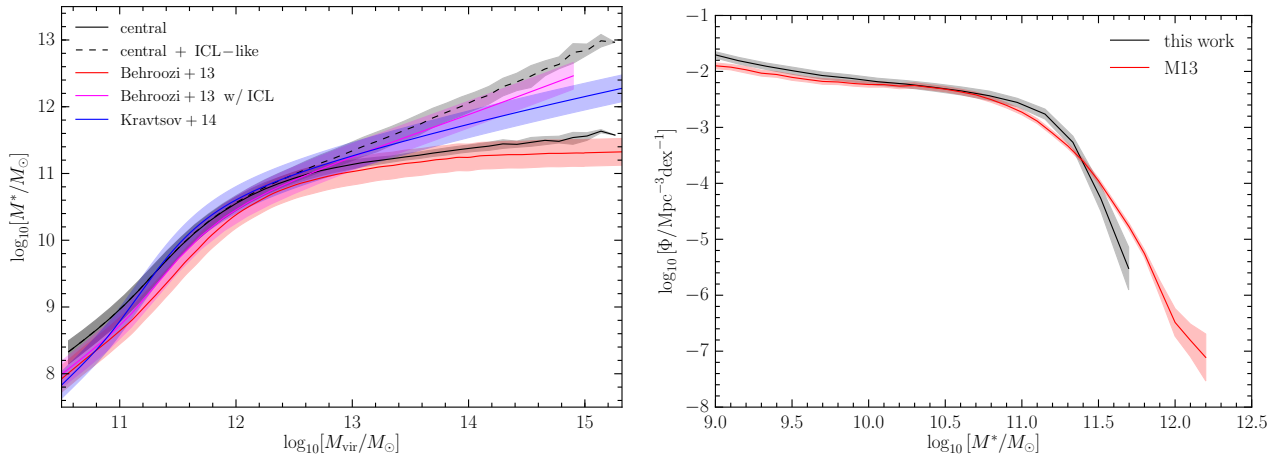
$$dM_{\text{vir},\text{cen}}/dt = \Delta M_{\text{vir},\text{cen}}/\Delta t \equiv \Delta M_{\text{vir}}/\Delta t. \quad (3)$$

where  $\Delta M_{\text{vir}}$  is the mass difference between two epochs along the most massive progenitor branch of the merger tree and  $\Delta t$  is the time spacing of the epochs.

In a merger, the growth in the mass of the halo arises from smooth accretion and the merging of the halos themselves. In this method, the halos which merge have themselves already been forming stars, so that if one counted all of the mass change, one would effectively double count the mass growth and thus stellar growth. Thus during mergers one should adjust the mass growth using a definition similar to the following

$$\Delta M_{\text{vir},\text{cen}} = \Delta M_{\text{vir}} - \frac{M_{\text{vir},\text{rest}}}{M_{\text{vir},\text{mmp}} + M_{\text{vir},\text{rest}}} M_{\text{vir},\text{desc}} \quad (4)$$

where  $M_{\text{vir},\text{mmp}}$  is the most massive progenitor halo mass and  $M_{\text{vir},\text{rest}}$  is the sum of the masses of all other progenitors.



**Figure 1.** The stellar mass-to-halo mass relation (left) and stellar mass function (right) of galaxies. In the left panel, the solid line shows the stellar mass-to-halo mass relation for the central galaxy in each halo, the dashed line shows the stellar mass in the central plus ICL-like variable for each halo, and the other lines show various constraints from low-redshift, mostly SDSS, data. The right panel shows the stellar mass function from this work and that from Moustakas et al. (2013).

Note that  $\Delta M_{\text{vir}} = M_{\text{vir,desc}} - M_{\text{vir,mmp}}$  so that this formula simply removes some fraction of the descendent halo’s mass that is presumed to be due to the merging of halos besides the most massive progenitor. This fraction is fixed to the fraction of total mass before the merger. Halo mergers are not additive (see, e.g., Kazantzidis et al. 2006), but as long as they are equally non-additive for all halos involved, this formula is roughly correct.

However, in this work, I use the Behroozi et al. (2013a) constraints which correct for the effect of mergers on the stellar mass assembly, but are phrased in terms of the median halo mass assembly along the most massive progenitor track in the merger tree. These mass accretion rates include the merging of smaller structures. Thus in this work the correction due to the second term of equation 4 is neglected.

### 3.1.2 Integrating the SMAD in Time

Once  $dM_*/dt$  is determined for each progenitor halo, the total stellar mass formed is simply  $dM_*/dt \times \Delta t$ . The next task is then to assign the stellar mass to the halos and evolve it along the tree. In this work, I track two different stellar mass variables. The first is the stellar mass associated with the central galaxy of each halo or subhalo. The second is an intra-cluster light-like (ICL-like) variable that holds stellar mass deposited by halos not along the most massive progenitor branch during mergers between three or more halos at any epoch.

Stellar mass is passed from progenitors to descendants according to the following rules:

**No Mergers:** For a halo that does not undergo a merger, the central stellar mass of the descendent is equal to the sum of the formed stellar mass plus all of the previously formed central stellar mass. The stellar mass in the ICL-like variable is simply passed to the descendent.

**Binary Mergers:** For mergers of just two halos, the central stellar mass of both halos is added together into the central stellar mass of the final halo, along with any formed stel-

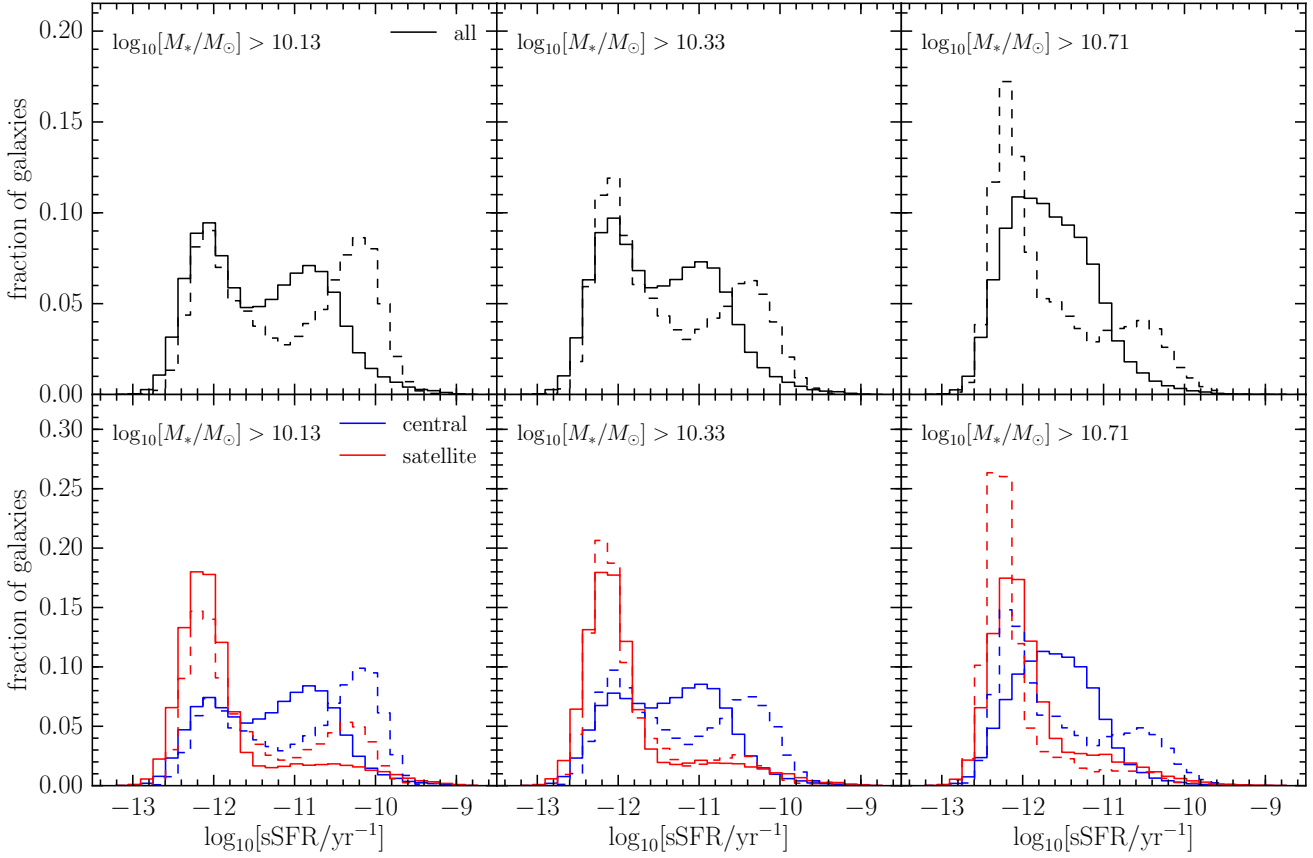
lar mass. The ICL-like variables from both halos are added together and then passed to the descendent.

**$N > 2$  Mergers:** For halos which have more than two progenitors, only the central stellar mass of the most massive progenitor halo is passed to the central stellar mass of the descendent halo, along with any formed stellar mass. The central and ICL-like stellar masses of the other progenitors are added to the ICL-like stellar mass of the descendent halo.

Note that in all cases, the newly formed stellar mass is always added to the central variable of the descendent halo.

### 3.1.3 Initial Conditions, Parameterization and Stellar Mass Loss

With the rules and definitions above, a set of halos at any epoch can be evolved to another epoch given the merger tree. In order to seed halos with stellar mass at the initial epoch, redshift four in this work, I use the Behroozi et al. (2013a) constraint on the  $M_* - M_{\text{vir,peak}}$  relation at  $z = 4$ . This choice is somewhat arbitrary and is motivated by the decreasing quantity galaxy population data above this epoch. Note that I make no adjustments to the Behroozi et al. (2013a) constraints, besides the minimal set of additional rules required to track the growth of stellar mass in halos in the central galaxy, in the ICL-like variable and the small amount of residual star formation seen in real galaxies. In particular, the details of the treatment of galaxy merging and the generation of intra-cluster light differ between Behroozi et al. (2013a) and this work. Thus there is no reason to expect exact quantitative agreement between the results of the two works. Finally, the stellar mass computed by this algorithm is the raw stellar mass assembled in the halos’ histories. In order to account for stellar mass loss, I use the Behroozi et al. (2013a) fit to single stellar population evolutionary tracks in order correct the stellar masses.



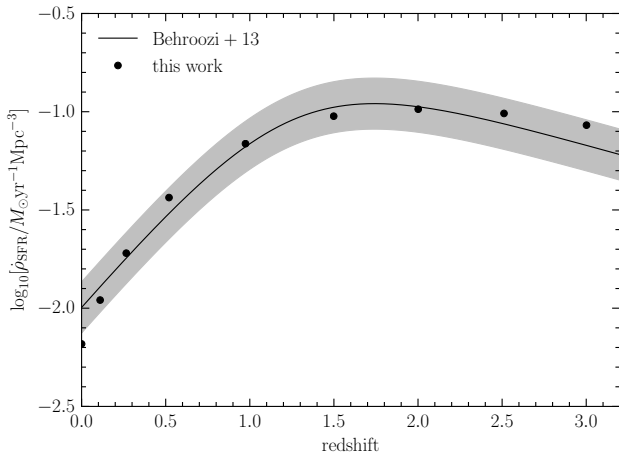
**Figure 2.** The distribution of specific star formation rates as a function of stellar mass. The dashed lines in the top panel show the results from the SDSS as represented by the CAM mock catalog of Watson et al. (2015). The bottom panels show results split into central (blue) and satellite (red) galaxies. The dashed lines in the bottom panel show the CAM prediction for centrals and satellites from Watson et al. (2015).

## 4 RESULTS

In this section, I explore the results of my simple model, illustrating its one and two-point statistics. I focus explicitly on local, low-redshift data from the SDSS. Note that while the model agrees qualitatively with data from the SDSS, the quantitative agreement is in some cases lacking. Future work will explore how such models can be directly fit to SDSS and higher redshift data. Before comparing to the data, note that because this model is not tuned to match the SDSS stellar mass function when integrated over the merger trees, there is an overall disagreement between the stellar masses in the SDSS and those from the model (see Figure 1). Thus to compare the distributions, I compare galaxies of the same number density as those in the SDSS. For the thresholds of  $\log_{10}[M_*/M_\odot] = \{9.8, 10.2, 10.6\}$  in the SDSS, I find that thresholds in my model of  $\log_{10}[M_*/M_\odot] = \{10.13, 10.33, 10.71\}$  match the number density in the SDSS. For each stellar mass threshold, the split in specific star formation rate in the mocks is determined by requiring the same fraction of galaxies to be either star forming or passive as determined by a threshold of  $\log_{10}[\text{sSFR}/\text{yr}^{-1}] = -11$  in the SDSS.

### 4.1 One-point Galaxy Statistics

Figure 1 shows the stellar mass-to-halo mass relation and the stellar mass function. The right panel of Figure 1 shows the stellar mass function from this work and that from Moustakas et al. (2013). The model error bars are from jackknife resampling of the simulation volume. While the agreement between the two is not perfect, the qualitative features, like the strongly decreased number density of massive galaxies, are correct. In the left panel of this figure, I show the results for the stellar mass of the central component (solid lines) and the stellar mass of the central plus the ICL-like component (dashed lines). I have also plotted the abundance matching relations from Behroozi et al. (2013a) and Kravtsov et al. (2014). I also show the Behroozi et al. (2013a) relation with the estimated ICL component included. The colored bands show the amount of scatter in abundance matching of  $\approx 0.2$  dex (see, e.g., Reddick et al. 2013). Note that the Behroozi et al. (2013a) and Kravtsov et al. (2014) relations differ in how the extended wings of light around BCGs are treated. The Kravtsov et al. (2014) relation is expected to include more of the light around BCGs and yields higher stellar masses for the most massive halos. Matching the aperture used to estimate the stellar mass in the data with the effective aperture implied by the central stellar mass vari-



**Figure 3.** The star formation rate density,  $\dot{\rho}_{\text{SFR}}$  as a function of redshift. The points show the results of this work. The line with grey band shows the constraints from data on  $\dot{\rho}_{\text{SFR}}$  used by Behroozi et al. (2013a) in their work.

able tracked above is difficult and beyond the scope of this work. However, my model does track the total stellar mass in both the central and ICL-like component, as inferred by Behroozi et al. (2013a), reasonably well. Finally, note that my predictions also produce a scatter in the stellar mass-to-halo mass relation, as shown by the grey band, that is smaller than, but close to the inferred scatter of  $\approx 0.2$  dex from abundance matching.

Figure 2 shows the distribution of specific star formation rates from my model. In the left, middle and right columns I show the distribution of specific star formation rate in cumulative bins of stellar mass (solid lines). The dashed lines show the results from the mock catalogs of Watson et al. (2015). The bottom panels show the results for centrals (blue) and satellites (red). I find that satellites in my model tend to be very quenched, possibly more so than in the SDSS group catalogs as described in Watson et al. (2015). Future extensions to the model based on the work of Wetzel et al. (2013), where satellite quenching is delayed, may be a useful in fixing this disagreement. Despite this difference the overall shift in the fraction of quenched galaxies follows the qualitative trends of the data. Note that the width and location of the quenched peak is directly determined by parameters in the model above. Also, my model adds negligible amount of scatter to the star forming sequence of galaxies. Thus the fact that the output star forming sequence has a scatter that is similar to the star forming peak in the SDSS indicates that the scatter in halo mass accretion rates contributions in a significantly to the scatter in star formation rates observed in the SDSS.

Finally, Figure 3 shows the density of star formation as a function of redshift, commonly called the Madau plot (Madau et al. 1998). The points show direct measurements of the star formation rate density from my model. The line and grey band are the summarized results from data used by Behroozi et al. (2013a). I find that my model successfully predicts the star formation rate density of the Universe. This agreement is expected since Behroozi et al. (2013a) fit the Madau plot as part of their analysis.

## 4.2 Two-point Clustering

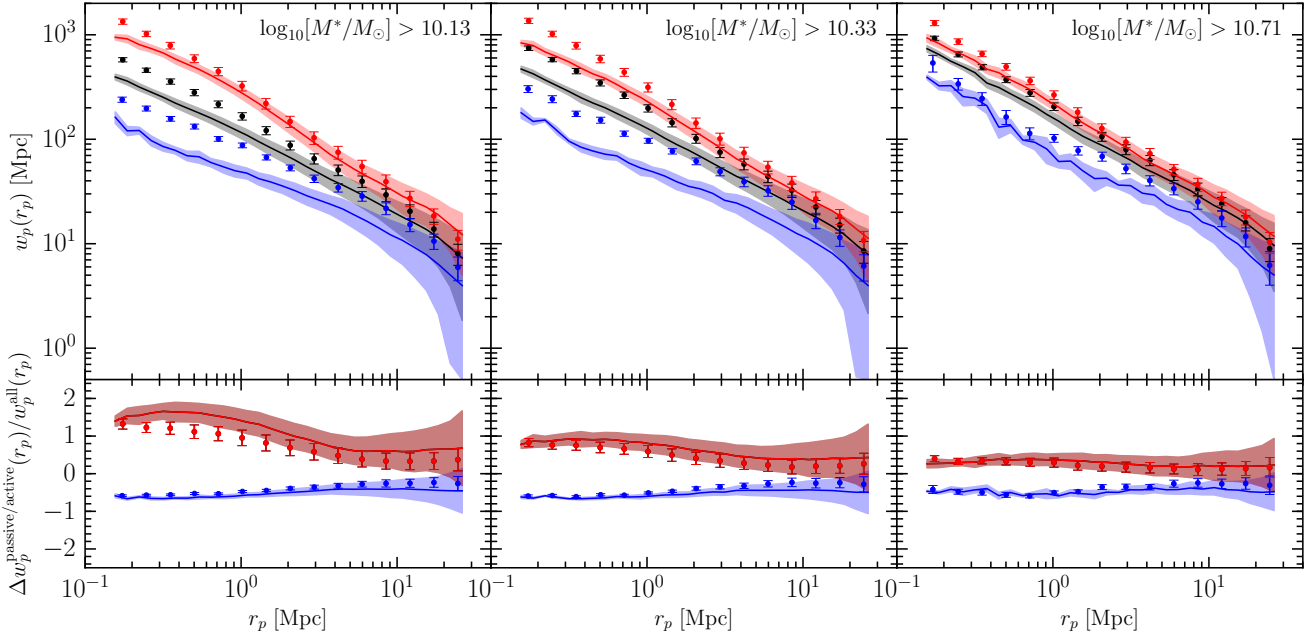
Figure 4 shows the star formation rate and stellar mass dependent two-point clustering for my model compared to recent measurements in the SDSS (Watson et al. 2015). The model error bars are from jackknife resampling of the simulation volume. I find that while the amplitude of the two-point clustering does not completely match the SDSS, the relative amplitude of the star forming and passive subsamples compared to the total clustering signals (shown in the bottom panels) does match the SDSS remarkably well. Note that Hearin et al. (2015) have shown that the mass accretion histories of halos are strongly influenced by their local environment through the tidal effects of other nearby halos. These effects generate the color dependent clustering in this model, directly through the mass accretion histories of individual halos in the merger trees. The partial success of this model indicates that assembly bias may play an important role in predicting the color dependent clustering seen in the SDSS, as proposed by Hearin & Watson (2013).

## 4.3 Stellar Mass Assembly Histories and Galaxy Colors

Finally, Figures 5 and 6 show the stellar mass assembly histories and predicted  $g-r$  colors of the galaxies in the SDSS bands. In Figure 5, I show the star formation histories in bins of stellar mass (*rows*) and for very quenched and very star forming galaxies (*columns*). The solid black line and grey bands show the median and middle 50% of the stellar mass assembly histories. The blue and red lines show the same data for centrals and satellites respectively. As expected, smaller galaxies tend to form the bulk of their stars later, whereas larger galaxies tend to form their stars quite early. This feature, noticed by Conroy & Wechsler (2009), arises naturally from the combination of the Behroozi et al. (2013a) constraints and the  $N$ -body merger trees (Neistein et al. 2006). Note also that the quenching of the star formation in massive halos is built directly in the Behroozi et al. (2013a) constraints. Finally, satellites tend to quench their star formation earlier than centrals, resulting in redder colors (see Figure 6) on average.

Figure 6 shows the SDSS  $g-r$  colors of galaxies in cumulative stellar mass bins. I have used the **FSPS** code from Conroy et al. (2009); Conroy & Gunn (2010) with solar metallicity and a Chabrier (2003) initial mass function.<sup>3</sup> I find that my model produces consistently red satellites at all stellar masses. It also produces some amount of color bimodality, with the amount of color bimodality increasing as the stellar mass decreases. However, there are serious disagreements between my simple model and SDSS data as represented by the (Watson et al. 2015) mock catalog (dashed lines in the figure). (Note the  $g-r$  color in the SDSS has been shifted down by 0.1 to account for an overall color difference between my mock and the SDSS colors.) In particular, low

<sup>3</sup> One must take care when running **FSPS** over the star formation histories. In particular, for the coarse star formation histories generated by the simulation merger tree, I found that accurate results could only be obtained by running the SPS assuming a piecewise constant star formation history with careful sub-integrations between each snapshot.



**Figure 4.** The two-point clustering as a function of stellar mass and specific star formation rate. The points show the measurements from Watson et al. (2015) of galaxies in the SDSS. The lines with bands show predictions from this work. Lines and points in red are quenched galaxies, in black are all galaxies, and in blue are active galaxies. the top panels show the projected correlation function and the bottom panels show the ratio of the correlation functions of active and passive galaxies to that of all galaxies.

stellar mass central galaxies appear to be too red. This feature is consistent with the distributions of sSFR's in Figure 2 where the star forming sequence as a whole appears to be a bit too quenched. Future work to directly fit the SMAD to the SDSS data should improve the model in these respects.

## 5 CONCLUSIONS

In this work, I have presented a general technique to model the galaxy-halo connect through the stellar mass assembly distribution (SMAD). In my implementation of this technique, the stellar mass assembly rates as constrained by Behroozi et al. (2013a) are used to construct an example SMAD. This model is then integrated directly over halo merger trees from  $N$ -body simulations, accounting for mergers of dark matter halos. I find that my model qualitatively reproduces the bimodal clustering of the halos as a function of their star formation rates. In this model, this bimodality is due to the correlations in the mass accretion histories of halos with their large scale environments (Hearin et al. 2015). My overall approach is very much like modern SAMs, but is directly statistical in nature. The SMAD parameterizes the ensemble of star formation rates at a fixed set of halo properties instead of attempting to predict the star formation rate correctly for any individual halo. Thus my model follows in the tradition of SHAM, CAM, HOD or CLF models which attempt to constrain the properties of galaxy formation directly from data as opposed to making *a priori* predictions.

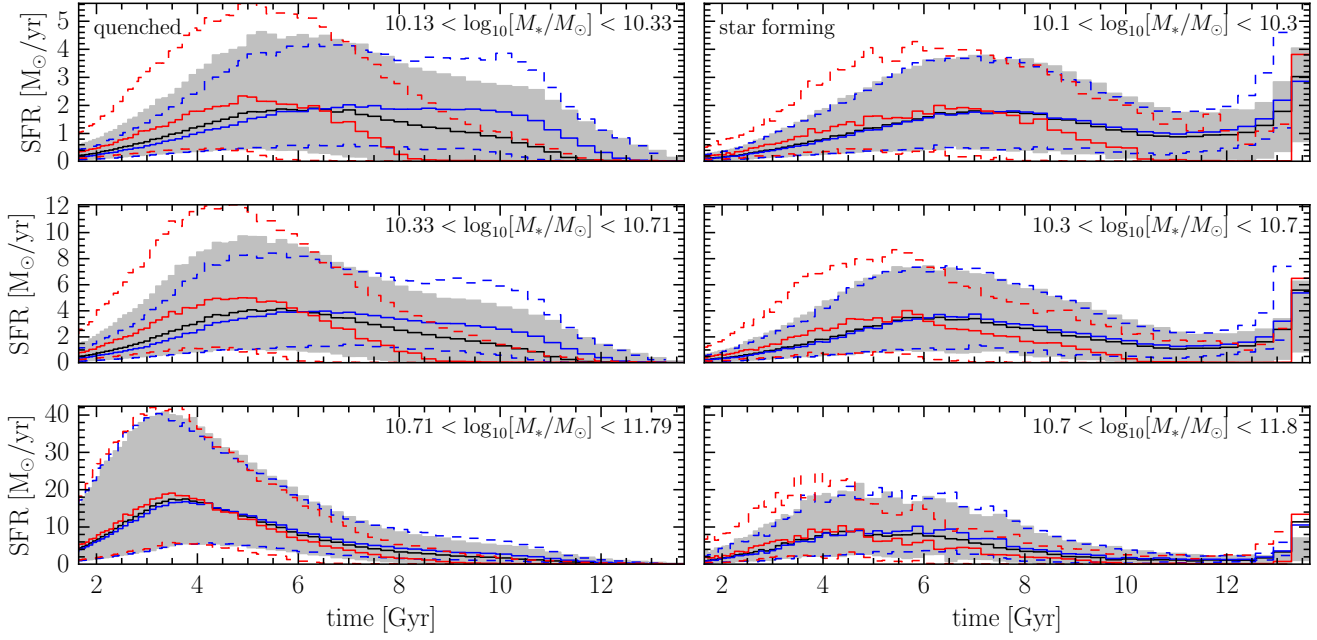
Future work which uses the framework outlined here to fit for the parameters which govern  $P(dM_*/dt|\mathbf{X}, a)$  using the one- and two-point statistics of the observed galaxy population should yield interesting constraints on the growth of

galaxies. Indeed, the general class of models discussed in this work can naturally reproduce observations of galactic conformity (e.g. Weinmann et al. 2006; Wang et al. 2010; Robotham et al. 2013; Kauffmann et al. 2013; Phillips et al. 2014; Knobel et al. 2015) and other potential signatures of assembly bias. These models also allow for natural extensions which can work to erase some of the assembly bias, such as adding correlated stochasticity in time to the stellar mass assembly rates. Simple extensions which, for example, randomly increase the star formation rates of galaxies in certain regimes, can account for other known effects like star forming BCGs in cooling flows, star bursts due to mergers, etc. Systematic changes in the stellar mass assembly rates of galaxies, say after they have reached their peak mass, may also be a useful modification (e.g. Wang et al. 2007; Wetzel et al. 2013). This model combined with stellar population synthesis models can be used to directly predict the colors and/or spectra of galaxies in broad band filters, as demonstrated in this work. Additionally, this model will provide predictions for the rates of SNe as a function of environment, stellar mass and redshift, as well as the mean star formation rate density as a function of time or any other statistic which can be predicted from an SPS code. These extensions and more detailed work fitting this model directly to data, such as clustering, lensing and redshift space distortions, will be the subject of future work.

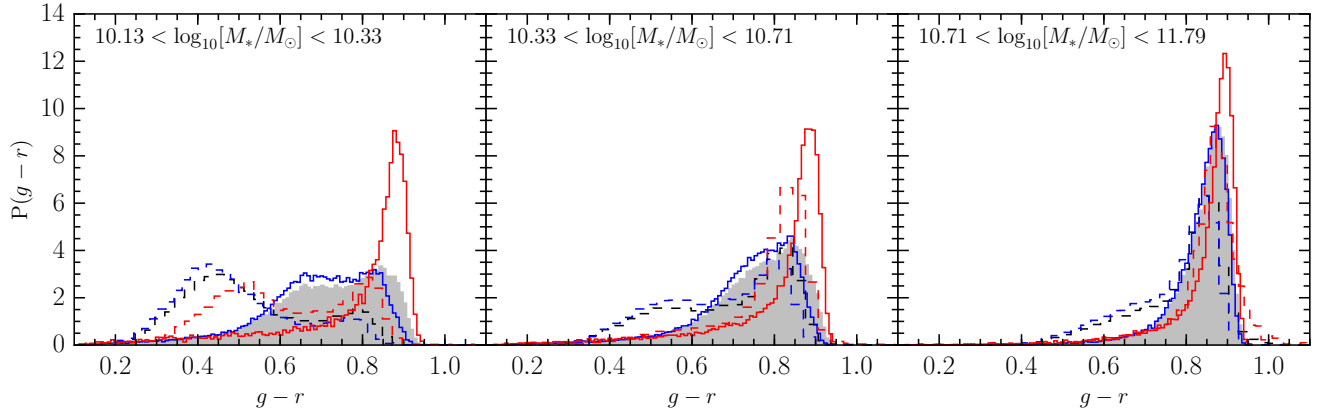
## ACKNOWLEDGEMENTS

I thank Andrew Hearin and Risa Wechsler for comments and encouragement during this work. I also thank Peter Behroozi, Doug Watson and Frank van den Bosch for useful com-





**Figure 5.** Star formation histories along the most massive progenitor branches. The solid lines show the median star formation history for all galaxies (*black*), central galaxies (*blue*) and satellite galaxies (*red*). The shaded grey band and dashed lines of the same color show the middle 50% deviations about the median histories.



**Figure 6.** SDSS  $g-r$  color distributions for galaxies in stellar mass bins. All galaxies are shown in the grey shaded region, centrals are in blue and satellites are in red. The dashed lines of the same colors show the distributions in the mock catalog of Watson et al. (2015) for all galaxies (*black*), centrals (*blue*) and satellites (*red*). Note the  $g-r$  color in the SDSS has been shifted down by 0.1 to account for an overall color difference between the colors in this work and the SDSS colors.

ments and discussion, and Yao-Yuan Mao for assistance with computing the projected correlation functions and simulation post-processing. This work made extensive use of the NASA Astrophysics Data System and the [arxiv.org](https://arxiv.org) preprint server.

## REFERENCES

- Abazajian K. N. et al., 2009, *ApJS*, 182, 543  
 Baugh C. M., 2006, *Reports on Progress in Physics*, 69, 3101  
 Behroozi P. S., Wechsler R. H., Conroy C., 2013a, *ApJ*, 770, 57  
 Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013b, *ApJ*, 762, 109  
 Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013c, *ApJ*, 763, 18  
 Berlind A. A., Weinberg D. H., 2002, *ApJ*, 575, 587  
 Bett P., Eke V., Frenk C. S., Jenkins A., Helly J., Navarro J., 2007, *MNRAS*, 376, 215  
 Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, *MNRAS*, 351, 1151



- Cacciato M., van den Bosch F. C., More S., Mo H., Yang X., 2013, *MNRAS*, 430, 767
- Chabrier G., 2003, *PASP*, 115, 763
- Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, *MNRAS*, 271, 781
- Conroy C., Gunn J. E., 2010, *ApJ*, 712, 833
- Conroy C., Gunn J. E., White M., 2009, *ApJ*, 699, 486
- Conroy C., Wechsler R. H., 2009, *ApJ*, 696, 620
- Conroy C., Wechsler R. H., Kravtsov A. V., 2006, *ApJ*, 647, 201
- Crocce M., Pueblas S., Scoccimarro R., 2006, *MNRAS*, 373, 369
- Diemer B., More S., Kravtsov A. V., 2013, *ApJ*, 766, 25
- Gao L., Springel V., White S. D. M., 2005, *MNRAS*, 363, L66
- Harker G., Cole S., Helly J., Frenk C., Jenkins A., 2006, *MNRAS*, 367, 1039
- Hearin A. P., Behroozi P. S., van den Bosch F. C., 2015, *arXiv:astro-ph/1504.05578*
- Hearin A. P., Watson D. F., 2013, *MNRAS*, 435, 1313
- Hearin A. P., Watson D. F., Becker M. R., Reyes R., Berlind A. A., Zentner A. R., 2014, *MNRAS*, 444, 729
- Kauffmann G. et al., 2003, *MNRAS*, 341, 33
- Kauffmann G., Li C., Zhang W., Weinmann S., 2013, *MNRAS*, 430, 1447
- Kazantzidis S., Zentner A. R., Kravtsov A. V., 2006, *ApJ*, 641, 647
- Knobel C., Lilly S. J., Woo J., Kovač K., 2015, *ApJ*, 800, 24
- Kravtsov A., Vikhlinin A., Meshscheryakov A., 2014, *arXiv:astro-ph/1401.7329*
- Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, *ApJ*, 609, 35
- Lewis A., Bridle S., 2002, *Phys. Rev. D*, 66, 103511
- Lilly S. J., Carollo C. M., Pipino A., Renzini A., Peng Y., 2013, *ApJ*, 772, 119
- Lu Z., Mo H. J., Lu Y., Katz N., Weinberg M. D., van den Bosch F. C., Yang X., 2014, *MNRAS*, 439, 1294
- Lu Z., Mo H. J., Lu Y., Katz N., Weinberg M. D., van den Bosch F. C., Yang X., 2015, *MNRAS*, 450, 1604
- Madau P., Pozzetti L., Dickinson M., 1998, *ApJ*, 498, 106
- Moster B. P., Somerville R. S., Maubetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, *ApJ*, 710, 903
- Moustakas J. et al., 2013, *ApJ*, 767, 50
- Mutch S. J., Croton D. J., Poole G. B., 2013, *MNRAS*, 435, 2445
- Neistein E., van den Bosch F. C., Dekel A., 2006, *MNRAS*, 372, 933
- Phillips J. I., Wheeler C., Boylan-Kolchin M., Bullock J. S., Cooper M. C., Tollerud E. J., 2014, *MNRAS*, 437, 1930
- Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, *ApJ*, 771, 30
- Robotham A. S. G. et al., 2013, *MNRAS*, 431, 167
- Salim S. et al., 2007, *The Astrophysical Journal Supplement Series*, 173, 267
- Sheth R. K., Tormen G., 2004, *MNRAS*, 350, 1385
- Somerville R. S., Davé R., 2014, *arXiv:astro-ph/1412.2712*
- Somerville R. S., Primack J. R., 1999, *MNRAS*, 310, 1087
- Springel V., 2005, *MNRAS*, 364, 1105
- van den Bosch F. C. et al., 2007, *MNRAS*, 376, 841
- Wang L., Li C., Kauffmann G., De Lucia G., 2007, *MNRAS*, 377, 1419
- Wang Y., Park C., Hwang H. S., Chen X., 2010, *ApJ*, 718, 762
- Watson D. F. et al., 2015, *MNRAS*, 446, 651
- Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, *ApJ*, 652, 71
- Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, *MNRAS*, 366, 2
- Wetzel A. R., Tinker J. L., Conroy C., van den Bosch F. C., 2013, *MNRAS*, 432, 336
- White S. D. M., Frenk C. S., 1991, *ApJ*, 379, 52
- York D. G. et al., 2000, *AJ*, 120, 1579
- Zehavi I. et al., 2011, *ApJ*, 736, 59